

of self-generated cognitive material as nonself in origin is thought to be a fundamental feature of psychotic phenomena [Frith and Done, 1988; Morrison and Baker, 2000]. The present study was designed to investigate the brain areas that are normally involved in differentiating self-generated speech from speech spoken by someone else. In a previous PET study, McGuire et al. [1996] presented participants with words which they read aloud. Participants heard the words spoken either in their own voice or by another person, and the feedback was either distorted by a pitch change or undistorted. They were required to decide whether the feedback was self or nonself in origin. Processing both nonself and distorted feedback was associated with activation of the lateral temporal cortices. However, as the accuracy of the responses was not recorded, the brain areas that were crucial to the correct attribution of speech were unclear. Fu et al. [2004] addressed this issue in a subsequent fMRI study of the same paradigm and found that the correct identification of self-generated speech was associated with greater activation of the lateral temporal cortex bilaterally than its misattribution to a nonself source. However, in both these studies participants were evaluating speech as they spoke aloud. Levelt [1983] proposes that self-monitoring can occur at three levels: 1) with the intention to speak, 2) when the intended output has been formulated but not yet articulated, and 3) at a sensory level following vocalisation when the speech is perceived. The discrimination of self from nonself speech may thus have involved the monitoring of verbal output with the intention to speak, in addition to the evaluation of the sensory feedback. In the present study, a similar paradigm was employed but participants were not required to speak; they simply listened to speech. Thus, the task primarily relied on the sensory evaluation of speech and did not involve cognitive self-monitoring of the intention to speak.

On the basis of the studies described above, we tested the hypothesis that processing speech in another person's voice and processing speech that was distorted would be associated with greater engagement of the lateral temporal cortices than listening to one's own speech or to undistorted speech. Following Allen et al. [2004], we predicated that participants would be particularly likely to misidentify their own speech as nonself when it was distorted, and that this misattribution would be associated with an attenuation of activation in the lateral temporal cortices.

SUBJECTS AND METHODS

Participants

Eleven right-handed, healthy, male volunteers ages 24–36 years (mean age = 28.09, SD = 3.91 years) participated. Subjects with a history of medical or psychiatric disorders, a drug or alcohol abuse problem, a family history of psychiatric disorder, or who were receiving medication were excluded. Their mean IQ (estimated with the National Adult Reading Test (NART) [Nelson and O'Connell, 1978]) was 115 (SD = 6.02, range 106–126). All participants spoke En-

glish as their first language. The study had Local Research Ethics Committee (LREC) approval. All participants had the nature of the experimental procedure explained fully and all gave informed consent.

Stimuli

Word lists

Eighty adjectives applicable to people were used (e.g., perfect, tall). All the words were mono- or bisyllabic with a Thorndike-Lorge frequency of >50 [Gilhooly and Logie, 1980], and were selected from lists used in a previous study by McGuire et al. [1996]. The emotional valence of the words had previously been rated by 40 healthy volunteers as either negative, positive, or neutral [see Johns et al., 2001]. In order to confirm that these ratings applied to the participants in the present study, the words were also rated by our participants following completion of the task. Words were considered negative if their mean rating was between -3 and -1, neutral between -0.9 and +1, and positive between +1.1 and +3. All participants' ratings fell within the expected ranges. Thus, the 80 words used consisted of 27 positive, 27 negative, and 26 neutral words.

The sets of words presented in each condition were balanced for the number of syllables (i.e., equal amounts of one- and two-syllable words), word frequency, and valence (equal amounts of positive, negative, and neutral words). Two lists of 40 words were generated; half of the participants received them in an AB sequence and half in BA sequence.

Auditory stimuli

The participants' speech was recorded on Cool Edit 2000 (for Windows). This software allowed the recordings to be normalised, pitch-shifted, and edited into 80 individual .wav files. The degree of pitch shift was -4 semitones, chosen because it made the speaker's voice harder to recognise without the speech becoming incomprehensible. A male researcher who was unknown to the participants recorded the words for the nonself condition (40 words in total).

Design

A 2×2 factorial design was used, with two sources of speech (self, alien) and two levels of distortion (0, -4 semitones). There were thus 20 words in each of four conditions (20 self-undistorted, 20 self-distorted, 20 alien-undistorted, and 20 alien-distorted). Thus, the experimental manipulations were the source of speech (self, alien) and distortion level (0, -4 semitones). Word valence was included as a factor in the behavioural analysis.

Words were presented in a nonself (alien) voice as well as the participant's voice to test whether any response bias was specific to self-generated words. The option to register an unsure response was included to avoid participants having to make a forced choice between a self or alien source when they were unsure. This increased the likelihood that when

they did make self and nonself attributions that they were confident about the source of the speech.

Procedure

Approximately 1 h before scanning participants were presented with a list of 80 words on a piece of paper and asked to read each one aloud in a clear voice at a rate of approximately one word per second. Participants were asked to read the words in a neutral voice so that they could not use prosody or intonation to facilitate source judgement later during the task. Participants read all 80 words even though half would subsequently be presented to them in another person's voice. These steps were taken to ensure that participants could not make judgements based on source information during the task.

Participants were not given the full task instructions at this stage but were told that they would have these words played back to them during the fMRI scan. They were not told to try and remember the words. Thus, the task was designed to rely on perceptual discrimination as opposed to source memory. Their speech was recorded by a computer. The experimenter then edited the set of recordings such that 40 of the words were replaced by a recording of the same word spoken in another person's voice and 40 of the words were pitch-shifted. The subsets of words that were replaced and pitch-shifted, respectively, were predesignated (allocated so that the subsets of replaced words were matched for word length and valence, etc.: see above). The same subsets of words were used for all participants.

Once they were in the scanner a standardised instruction script was read to the participants. Participants were told that if they thought the speech they heard was their own they were to press a button corresponding to "self." If unsure of its identity they were to press a button corresponding to "unsure," and if they thought the speech belonged to someone else they were to press a button corresponding to "Other." Participants did this using their right index finger. When they had indicated that they understood these instructions the words were presented one at a time via headphones: the words were not presented visually. The inter-stimulus interval was varied between 4 and 12 s to counteract possible habituation effects. Throughout the task participants were asked to look at a fixation cross, beneath the words "SELF," "UNSURE," and "OTHER" were continuously displayed. When participants made a response the corresponding word changed colour on the computer screen, confirming to the participant and the experimenter that the responses had been registered. A computer recorded the type of response.

Image Acquisition

Images were acquired in a 1.5 T Magnet (Signa LX-GE, Milwaukee, WI) using a compressed gradient echo [Edmister et al., 1999], echo planar image acquisition [Hall et al., 1999], with a TR of 1.2 s (0.8 s of silence), flip angle 80°, TE 40 ms, 64 × 64 pixels, field of view of 200 mm, slice thickness 7 mm, and interslice gap 0.7 mm (voxel size: 3.125 × 3.125

× 7 mm). In all, 482 image volumes were acquired in two runs of 6 min. Of these 482 image volumes, 80 were experimental events, whilst the remainder were null events used for baseline contrast. Each whole brain volume consisted of 14 axial slices parallel to the AC-PC line.

Stimuli were presented in random order in an event-related design, with a variable interstimulus interval (4–12 s) following a non-Gaussian random distribution (Poisson function peaking at 7 s) individually set for each condition [Dale, 1999]. Image acquisition and stimulus presentation were synchronised via a TTL pulse from the scanner to the computer used to present the stimuli and record the behaviour. The compressed acquisition permitted presentation of each word in the absence of acoustic scanner noise. Each response time was locked to the beginning of the word presentation.

Behavioural Analysis

The mean proportions of correct, unsure, and misattributed trials were calculated. Analysis of variance (ANOVA) was conducted for misattribution errors. Not all the error data met the assumption of normality, but repeated measures ANOVA was used since the normality assumption can be violated without significantly affecting the results [Howell, 1992]. However, to confirm the results of the parametric statistical tests, a Wilcoxon Signed Rank nonparametric test for related samples was also conducted. The within-subject factors were source of speech (self, alien), level of distortion (0, −4 semitones), and word valence (positive, negative, neutral).

Individual Analysis

The data were realigned [Bullmore et al., 1999] to minimise motion-related artefacts and smoothed using a Gaussian filter (FWHM 7.2 mm). As this was a multirun experiment, registration was achieved by concatenating the second functional to the average of the first functional run using XBAM v2 [Brammer et al., 1997]. A time-series analysis using Gamma variate functions (peak responses at 4 s and 8 s) was used to model the BOLD response to the task. Each experimental condition was convolved separately with the 4-s and 8-s Poisson functions to yield two models of the expected haemodynamic response. The weighted sum of these two convolutions that gave the best fit to the time series at each voxel was then computed. This weighted sum effectively allows voxel-wise variability in time to peak haemodynamic response. In order to constrain the possible range of fits to physiologically plausible BOLD responses, the fitting procedure suggested by Friman et al. [2003] was adopted. A goodness of fit statistic, the SSQ ratio, was then computed at each voxel. This was the ratio of the sum of squares of deviations from the mean intensity value due to the model (fitted time series) divided by the sum of squares due to the residuals (original time series minus model time series). The percentage change in BOLD signal at each voxel was also calculated: [(fitmax − fitmin)/mean signal intensity] * 100, where fitmax and fitmin were the maximum and minimum values of the fitted response for the time series in question.

TABLE I. Correct, misattributed, and unsure responses by condition.

	Self	Self Distorted	Alien	Alien distorted
Correct responses	17.00 (2.82)	6.00 (4.73)	15.00 (5.40)	12.50 (4.74)
Misattributions	1.6 (2.10)	7.80 (7.13)	3.80 (4.80)	3.50 (5.10)
Unsure responses	1.4 (1.50)	6.20 (5.72)	1.20 (1.61)	4.00 (3.13)

Values are expressed as mean (SD).

Table data does not include null responses.

In order to sample the distribution of SSQ ratio under the null hypothesis that the observed values of SSQ ratio were not determined by the experimental design, the time series at each voxel was permuted using a wavelet-based resampling method [Bullmore et al., 2001; Breakspear et al., 2004]. This process was repeated 20 times at each voxel and the data were combined over all voxels, resulting in 20 permuted parametric maps of SSQ ratio in each plane for each. The same permutation strategy was applied at each voxel to preserve spatial correlational structure in the data during randomisation. Combining the randomised data over all voxels yielded the distribution of SSQ ratio under the null hypothesis.

Group Mapping

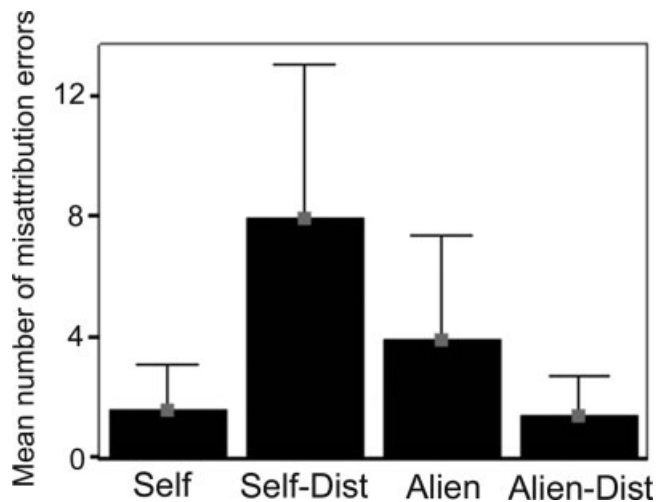
The observed and randomised SSQ ratio maps from each were transformed into a standard stereotactic space by a two-stage process involving a rigid body transformation of the data onto a high-resolution inversion recovery image from the same, followed by an affine transformation on to a Talairach template [Brammer et al., 1997]. A generic brain activation map (GBAM) was then produced for each experimental condition. The median observed SSQ ratio for all participants at each voxel was tested at each intracerebral voxel against a critical value of the permutation distribution for median SSQ ratio ascertained from the spatially transformed wavelet-permuted data [Brammer et al., 1997]. In order to increase sensitivity and reduce the number of comparisons, hypothesis testing was carried out at the cluster level [Bullmore et al., 1999]. This estimated the probability of occurrence of clusters under the null hypothesis using the distribution of median SSQ ratios computed as above. Image-wise expectation of the number of false-positive clusters under the null hypothesis was set for each analysis at <1 .

Repeated Measures Contrasts

ANOVA was carried out on the SSQ ratio maps in standard space by computing the difference in median SSQ ratio between groups at each voxel. The probability of this difference under the null hypothesis was inferred by reference to the null distribution obtained by repeated random permutation of group membership and recomputation of the difference in median SSQ ratios between the two groups obtained from the resampling process. Cluster-level maps were then obtained as described above.

The experimental conditions were defined according to the source of the speech (self or alien) and the level of distortion (undistorted or distorted). The data were analysed using a nonparametric within- 2×2 factorial design [Bullmore et al., 1999]. This design included the source of speech and the level of the distortion as factors. This revealed the main effects of speech source (independent of distortion) and of distortion (independent of source) and their interaction. The main effects analysis for word valence was conducted by comparing all emotional words (positive and negative in a single category) with all neutral words across all experimental conditions.

A further analysis examined the main effect of the accuracy of attribution (correct vs. incorrect) for all conditions. Each participant's behavioural responses were used to categorise each event as associated with a correct attribution, an unsure response, or a misattribution. As we were particularly interested in the correlates of misattributions of self-generated speech, we compared trials associated with correct responses and misattributions in the self-distorted speech condition, as this condition is particularly associated with external misattributions. Trials associated with unsure responses were excluded from

**Figure 1.**

Mean number of misattribution errors by condition.

TABLE II. Coordinates of foci of activation whilst listening to and evaluating the source of speech

Cerebral region	Side	<i>x, y, z</i>	Size	SSQ ratio	BA
Superior temporal gyrus	R	46, -19, -2	136	0.03	22
Superior temporal gyrus	L	-46, -22, 9	55	0.08	42
Middle temporal gyrus	L	-49, -28, -2	15	0.02	21
Fusiform gyrus	L	-35, -53, -13	65	0.01	20
Inferior frontal gyrus	R	35, 25, 4	121	0.03	45
Inferior frontal gyrus	L	-40, 19, 15	12	0.02	45
Anterior cingulate gyrus	R	3, 14, 26	42	0.03	24
Posterior cingulate gyrus	R	6, -58, 15	17	0.02	23
Precuneus	R	6, -61, 26	46	0.05	31
Cerebellum	L	-1, -39, -7	35	0.01	—
Brainstem	L	-2, -18, -7	12	0.02	—

Coordinates refer to the stereotactic space as defined in the atlas of Talairach and Tournoux (1988).

P cluster < 0.001. All clusters significant at a type I error expectancy of < 1 per brain (see Table I).

BA, Brodmann area.

this analysis. Maps of the difference in the effect size of the BOLD response associated with correct and incorrect attributions were generated. The effect size statistic was used because the numbers of trials associated with correct and incorrect responses was not equal and it is relatively insensitive to differences in the number of responses per condition (although the accuracy with which it can be estimated is a function of the number of responses averaged in the model-fitting procedure). Use of the effect size statistic also avoids the possibility that differences in BOLD response could reflect changes in the denominator of the statistic (noise) rather than signal, as can occur when using standardised statistics such as *F*, *t* or SSQ ratio.

RESULTS

Behavioural Results

The numbers of correct, unsure and error trials are displayed in Table I and Figure 1.

Misattribution Errors

There was a trend for a main effect of source of speech ($F = 3.69$, $df = 1,10$, $P = 0.08$), while the main effect of distortion ($F = 2.69$, $df = 1,10$, $P = 0.14$) was nonsignificant. There was a significant interaction between source of speech and distortion ($F = 5.59$, $df = 1,10$, $P = 0.04$). A series of pairwise *t*-tests revealed that when processing their own distorted speech participants made significantly greater misattribution (vs. self-undist: $t = -2.49$, $df = 10$, $P = 0.03$, vs. alien-dist: $t = 2.78$, $df = 10$, $P = 0.02$). A nonparametric Wilcoxon Signed Rank Test confirmed these results ($z = 2.19$, $P = 0.028$, and $z = 2.89$, $P = 0.03$). The main effect for word valence was nonsignificant ($F = 1.91$, $df = 2,18$, $P = 0.17$). There were no significant interactions between word valence and the source of speech.

fMRI Data

Activation associated with evaluating the source of auditory speech

Performance of the task across all conditions (independent of performance) was associated with bilateral activation in the superior temporal, and the inferior and middle frontal gyri, and activation in the anterior and posterior cingulate gyri, precuneus, cerebellum, and brain stem and in the left middle temporal and fusiform gyri (Table II).

Comparisons Between Conditions

Self-generated vs. alien speech

Processing self-generated words was associated with more activation than listening to words in an alien voice in the left inferior frontal gyrus and the anterior cingulate gyrus (Fig. 2a; Table III). Conversely, processing words in another person's voice was associated with greater activation in the superior temporal and lingual gyri bilaterally (Fig. 2b, Table III).

Distorted speech vs. undistorted speech

Processing distorted words was associated with more activation than undistorted words in the inferior frontal and anterior cingulate gyri bilaterally (Table III, Fig. 2c). Conversely, listening to undistorted speech was associated with relatively greater engagement of the left middle temporal gyrus (Table III, Fig. 2d).

Valence

Main effect for emotional vs. neutral words

Emotional words were associated with more activation than neutral words in the left middle temporal gyrus, supramarginal gyrus, and precuneus.

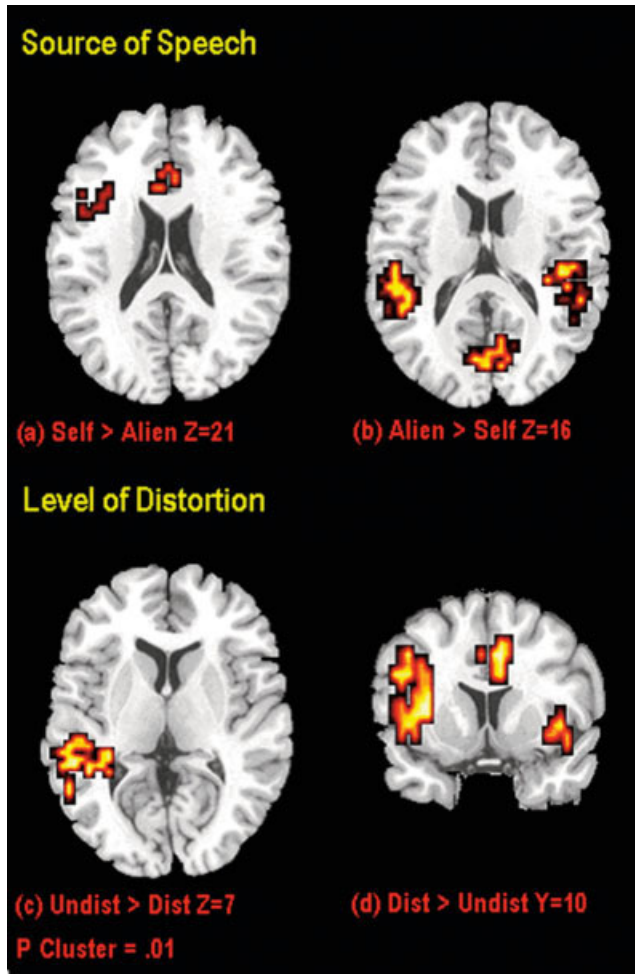


Figure 2.

a–d: Brain activation maps of the main effects contrasts for source of speech (self speech vs. alien speech) and the level of distortion (nondistorted vs. distorted speech). The left side of the brain is shown on the left side of the images. The level of the axial and coronal sections are indicated by their Z and Y coordinates in mm, respectively.

Interaction

There was an interaction between the source of speech and the level of distortion in the left superior temporal gyrus. Examination of the SSQ ratios contributing to this cluster revealed that in this region distortion had little effect on the response to speech in an alien voice but markedly attenuated the response when the words had been self-generated (Fig. 3).

Misattributions vs. Correct Responses

Relative to correct responses, misattributions (across all conditions) were associated with reduced activation in the lateral temporal and inferior frontal cortex bilaterally, the anterior cingulate gyrus, and the right lingual gyrus. There were no regions that were more active in associa-

tion with misattributions than correct responses. When the analysis was restricted to the self-distorted speech condition correct responses were associated with greater activation than misattributions in the anterior cingulate and left inferior frontal cortex (see Fig. 4) and Table IV.

DISCUSSION

The present study used functional MRI to study the neural correlates of making self/nonself judgements about the source of prerecorded words. The task was made more difficult by manipulating the pitch of the speech such that it sounded distorted. We were thus able to examine the brain areas involved in evaluating auditory speech, and assess how its source and acoustic quality affected activity in this network. We also categorised the neural response to each word according to the accuracy of the self/nonself attribution and compared activity when participants misidentified the source of the speech with that when they correctly identified its origin. The potentially confounding effects of acoustic scanner noise were minimised by acquiring images with a compressed sequence that permitted the presentation of each word during a brief period of silence.

Misattribution errors were most frequent when distortion was applied to recordings of the participants' own speech, as opposed to another person's speech. Thus, when the source of the speech was ambiguous there was an increased likelihood that self-speech would be misattributed to an external source. Previous studies have shown that, although this bias is particularly evident in patients with schizophrenia, it is also evident (albeit in an attenuated form) in healthy volunteers [Allen et al., 2004; Johns et al., 2001].

As expected on the basis of previous studies examining voice processing, listening to and evaluating words (independent of source, distortion, and accuracy of attribution) was associated with activation in the lateral temporal [Binder et al., 2000] and inferior frontal cortex [Vouloumanos et al., 2001]. More precisely, Belin et al. [2000] described bilateral foci along the upper bank of the anterior and posterior superior temporal sulcus as well as in the left middle temporal gyrus. In the present study activation in the left superior temporal sulcus was also in its upper bank, with the focus midway along the sulcus. There was also engagement of the anterior cingulate gyrus, fusiform gyrus, cerebellar cortex, and precuneus (a region also reported in previous voice processing studies: see Belin et al. [2000]).

When participants processed words spoken in their own voice as opposed to another's, there was increased engagement of the left inferior frontal and the right anterior cingulate gyrus. The finding that the cingulate gyrus was more activated by self-generated speech is consistent with data from studies of a task which was similar to that used in the present study, save that participants evaluated words that they read aloud [Fu et al., 2001] and more general evidence that cortical midline structures are involved in self-process-

TABLE III. Coordinates of foci of activation for between condition comparisons*

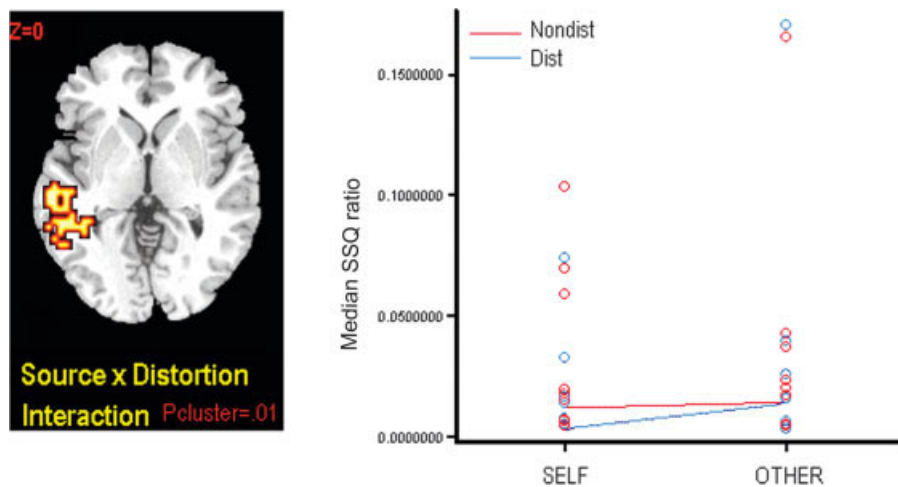
Cerebral region	Side	x, y, z	Cluster	SSQ ratio	BA
Source					
Self > alien					
Inferior frontal gyrus	L	-36, 20, -2	69	0.01	47
Anterior cingulate gyrus	R	3, 22, 25	67	0.01	24
Alien > self					
Middle temporal gyrus	R	49, -26, -3	82	0.02	21
Superior temporal sulcus (junction of MTG and STG)	L	-45, -29, -1	78	0.006	21/22
Lingual gyrus	L	-6, -67, -3	120	0.01	18
Dist level					
Dist > undist					
Inferior frontal gyrus	R	42, 19, 0	67	0.01	47
Inferior frontal gyrus	L	-34, 20, 0	153	0.02	47
Anterior cingulate gyrus	R	4, 22, 28	107	0.002	32
Undist > dist					
Middle temporal gyrus Valence	L	-47, -33, -2	145	0.004	21
Valence					
Emotional > neutral					
Middle temporal gyrus	L	-51, -41, -2	45	0.003	22
Supramarginal gyrus	L	-40, -37, 31	15	0.001	40
Precuneus	L	-22, -74, 26	9	0.001	40
Source \times Dist interaction					
Superior temporal sulcus	L	-47, -22, -7	96	0.015	21/22

* P cluster < 0.01.

Coordinates refer to the voxel with the maximum SSQ ratio in each cluster in stereotactic space as defined in the atlas of Talairach and Tournoux (1988).

ing [Northoff and Bermpohl, 2004]. The specific executive processes engaged by self as opposed to nonself stimuli are unclear. It is unlikely to be simply an effect of increased attention to the auditory stimuli, as this normally increases (rather than decreases) temporal activation [Woodruff et al., 1997]. One possibility is that activation of the inferior frontal gyrus may be associated with the processing of a nonphonetic “voice” aspect of speech [Vouloumanos et al., 2001]

In contrast, processing nonself words was associated with greater engagement of the superior temporal gyri. It is possible that this reflected a response to the different acoustic characteristics of nonself speech. Self-generated speech would be more familiar than nonself speech and may place less of a demand on acoustic processing. However, the greater engagement of inferior frontal and anterior cingulate cortex when listening to self-generated words may have

**Figure 3.**

Brain activation maps and median SSQ plots for interaction between the source of speech and distortion in the left superior temporal gyrus.

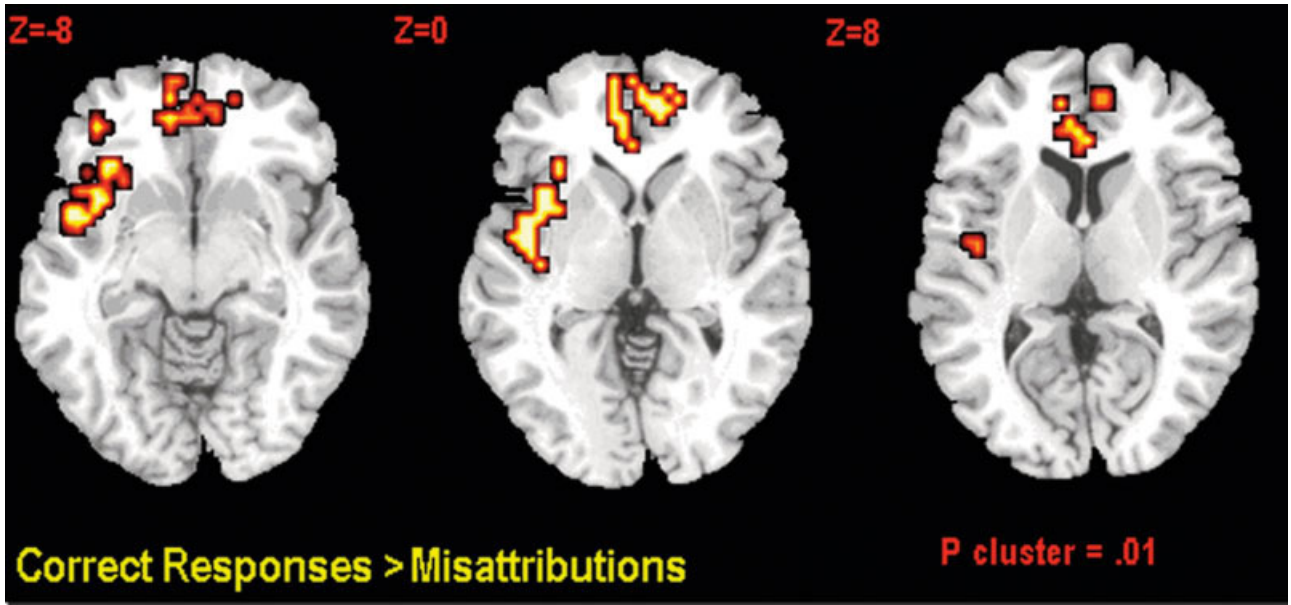


Figure 4.

Effect size maps of areas more activated in association with correct responses than misattributions when processing distorted self speech.

modulated the response in the temporal cortex, such that it was less active when processing self-generated words. The similarity of the effects of distorted speech on regional activation is consistent with this interpretation. Thus, processing distorted speech also led to greater activation in the same parts of the inferior frontal and anterior cingulate gyri as self-generated speech (although in this case the inferior frontal changes were bilateral), and a relative attenuation of activation in the (left) temporal cortex. In this case these effects may have reflected a greater engagement of executive functions such as response selection [Carter et al., 1998] and/or the process-

ing of response conflict [van Veen et al., 2001] when the pitch change made identification of the speaker more difficult. The inferior frontal and anterior cingulate cortex are strongly interconnected with the lateral temporal cortex [Petrides and Pandya, 1988], and there is electrophysiological evidence in nonhuman primates and humans that activity in the former regions during linguistic processing can suppress temporal cortical activity [Ford and Mathalon, 2004; Muller-Preuss and Ploog, 1981]. However, the present study did not explicitly examine the modulatory effects of prefrontal/cingulate regions on the auditory cortex and this issue would need to be addressed

TABLE IV. Coordinates of foci of activation for response conditions main effect and interactions*

Cerebral region	Side	<i>x, y, z</i>	Cluster	BA
Correct > misattributions				
Middle temporal gyrus	L	-47, -18, -10	284	21
Middle temporal gyrus	R	50, -11, -12	192	21
Lingual gyrus	R	15, -56, -8	126	19
Inferior frontal gyrus	R	41, 17, 13	74	9
Cingulate gyrus	R	7, 11, 31	116	24
Inferior frontal gyrus	L	-36, 26, -12	67	47
Distorted self-speech				
Correct > misattributions				
Anterior cingulate	R	4, 37, 4	56	24
Inferior frontal/insula	L	-32, 30, -18	88	47

* *P* cluster < 0.01.

Coordinates refer to the voxel with the maximum percentage signal change (effect size) in each cluster in stereotactic space as defined in the atlas of Talairach and Tournoux (1988).

in further studies of the effective connectivity between these regions.

An interaction between the effects of the source of speech and distortion was evident in the left superior temporal gyrus. Activity in this region was sensitive to the source of the perceived speech. Moreover, when the acoustic quality of the speech was degraded its response was attenuated. However, the effect of distortion in this region was specific for words which were self-generated. The selectivity of this effect is difficult to attribute to differences in the sensory features of the stimuli and again may reflect prefrontal/anterior cingulate modulation of auditory sensory regions. These effects may have been evident in the left temporal cortex (as opposed to both the left and the right) because it is particularly associated with auditory verbal processing [Scott et al., 2000].

Activation in the left temporal cortex, supramarginal gyrus, and precuneus was significantly greater for emotional words when compared to neutral words. This is consistent with previous reports of increased activation in the superior and middle temporal regions in response to pleasant and unpleasant words compared to neutral words. [Maddock et al., 2003; Tabert et al., 2001].

In a previous behavioural study [Allen et al., 2004], we found that subjects were particularly likely to misidentify the source of the words when they processed self-generated speech that was distorted. The same is true when patients with schizophrenia are tested with this paradigm, although the magnitude of this effect is even greater [Allen et al., 2004]. We examined the neural correlates of this verbal misattribution by comparing activation when participants misattributed their own distorted speech to another person with activation during correct "self" attributions. The misattribution of self-generated speech was associated with attenuated engagement of the left inferior frontal and the anterior cingulate cortex. This suggests that accurate source identification of speech may be facilitated by the engagement of these areas. Failure to engage these areas might therefore contribute to the misidentification of self-generated verbal material as "alien," thought to be the fundamental deficit underlying auditory verbal hallucinations [Frith and Done, 1988].

The involvement of inferior frontal, cingulate, and temporal cortex in this process is consistent with neuroimaging studies in schizophrenia, which indicate that these regions are critically implicated in the pathophysiology of auditory hallucinations [McGuire et al., 1993; Shergill et al., 2000, 2003]. However, the paradigm used in the present study involved the evaluation of external rather than inner speech, which is more relevant to verbal hallucinations. Nevertheless, it is possible that the same brain regions are involved in differentiating self from nonself speech, whether the speech is processed externally or internally.

Because there was some variation in the frequency and length of the words used as stimuli it is possible that this may have influenced the activation we observed. For example, Kronbichler et al. [2004] reported that word frequency

can modulate activation in frontal areas. However, because the frequency and length of the words was matched across the experimental conditions, any effects of these factors will have been minimised. The short time between when participants recorded the words and undertook the task could have allowed participants to use a memory of how they originally spoke the words to help them make judgements about their source. However, because participants were required to read *all* the words that they subsequently heard during the task, including those that would be presented in the alien voice, any memory effect would have applied equally to words presented in a self and a nonself voice. Thus, the participant's judgements about the identity of the voice depended on an assessment of its acoustic qualities rather than whether they had read the words beforehand.

Overall, the data suggest that manipulating the source and acoustic quality of external speech particularly affects the engagement of the inferior frontal, anterior cingulate, and temporal cortices, and that there may be a reciprocal relationship between activity in these regions which reflects their interconnections. The latter could be examined through an analysis of functional and effective connectivity [Honey et al., 2003]. The putative influence of anterior cingulate and inferior frontal cortex on temporal activation in this context may reflect the involvement of executive processes that are involved in the modulation of auditory verbal processing.

REFERENCES

- Allen PP, Johns LC, Fu CH, Broome MR, Vythelingum GN, McGuire PK (2004): Misattribution of external speech in patients with hallucinations and delusions. *Schizophr Res* 69:277–287.
- Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000): Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex* 10:512–528.
- Brammer MJ, Bullmore ET, Simmons A, Williams SC, Grasby PM, Howard RJ, Woodruff PW, Rabe-Hesketh S (1997): Generic brain activation mapping in functional magnetic resonance imaging: a nonparametric approach. *Magn Reson Imag* 15:763–770.
- Breakspear M, Brammer MJ, Bullmore ET, Das P, Williams LM (2004): Spatiotemporal wavelet resampling for functional neuroimaging data. *Hum Brain Mapp* 23:1–25.
- Bullmore ET, Brammer MJ, Rabe-Hesketh S, Curtis VA, Morris RG, Williams SC, Sharma T, McGuire PK (1999): Methods for diagnosis and treatment of stimulus-correlated motion in generic brain activation studies using fMRI. *Hum Brain Mapp* 7:38–48.
- Cahill C, Silbersweig D, Frith CD (1996): Psychotic experiences induced in deluded patients using distorted auditory feedback. *Cogn Neuropsychiatry* 1:201–211.
- Carter CS, Braver TS, Barch DM, Botvinick MM, Noll D, Cohen JD (1998): Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* 280:747–749.
- Dale AM (1999): Optimal experimental design for event-related fMRI. *Hum Brain Mapp* 8:109–114.
- Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM (1999): Improved auditory cortex imaging using clustered volume acquisitions. *Hum Brain Mapp* 7:89–97.
- Ford JM, Mathalon DH (2004): Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *J Psychiatr Res* 38:37–46.

- Friman O, Borga M, Lundberg P, Knutsson H (2003): Adaptive analysis of fMRI data. *Neuroimage* 19:837–845.
- Frith CD, Done DJ (1988): Towards a neuropsychology of schizophrenia. *Br J Psychiatry* 153:437–443.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999): “Sparse” temporal sampling in auditory fMRI. *Hum Brain Mapp* 7:213–223.
- Honey GD, Sharma T, Suckling J, Giampietro V, Soni W, Williams SC, Bullmore ET (2003): The functional neuroanatomy of schizophrenic subsyndromes. *Psychol Med* 33:1007–1018.
- Johns LC, Rossell S, Frith C, Ahmad F, Hemsley D, Kuipers E, McGuire PK (2001): Verbal self-monitoring and auditory verbal hallucinations in patients with schizophrenia. *Psychol Med* 31:705–715.
- McGuire PK, Shah GM, Murray RM (1993): Increased blood flow in Broca’s area during auditory hallucinations in schizophrenia. *Lancet* 342:703–706.
- McGuire PK, Silbersweig DA, Frith CD (1996): Functional neuroanatomy of verbal self-monitoring. *Brain* 119:907–917.
- Morrison AP, Baker CA (2000): Intrusive thoughts and auditory hallucinations: a comparative study of intrusions in psychosis. *Behav Res Ther* 38:1097–1106.
- Muller-Preuss P, Ploog D (1981): Inhibition of auditory cortical neurons during phonation. *Brain Res* 215:61–76.
- Nelson HE, O’Connell A (1978): Dementia: the estimation of pre-morbid intelligence levels using the New Adult Reading Test. *Cortex* 14:234–244.
- Northoff G, Bermpohl F (2004): Cortical midline structures and the self. *Trends Cogn Sci* 8:102–107.
- Petrides M, Pandya DN (1988): Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *J Comp Neurol* 273:52–66.
- Scott SK, Blank CC, Rosen S, Wise RJ (2000): Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123:2400–2406.
- Shergill SS, Bullmore E, Simmons A, Murray R, McGuire P (2000): Functional anatomy of auditory verbal imagery in schizophrenic patients with auditory hallucinations. *Am J Psychiatry* 157:1691–1693.
- Shergill SS, Brammer MJ, Fukuda R, Williams SC, Murray RM, McGuire PK (2003): Engagement of brain areas implicated in processing inner speech in people with auditory hallucinations. *Br J Psychiatry* 182:525–531.
- van Veen V, Cohen JD, Botvinick MM, Stenger VA, Carter CS (2001): Anterior cingulate cortex, conflict monitoring, and levels of processing. *Neuroimage* 14:1302–1308.
- Vouloumanos A, Kiehl KA, Werker JF, Liddle PF (2001): Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *J Cogn Neurosci* 13:994–1005.
- Woodruff PW, Wright IC, Bullmore ET, Brammer M, Howard RJ, Williams SC, Shapleske J, Rossell S, David AS, McGuire PK, et al. (1997): Auditory hallucinations and the temporal cortical response to speech in schizophrenia: a functional magnetic resonance imaging study. *Am J Psychiatry* 154:1676–1682.